Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 54 (2012) 470 – 482

Procedia
Social and Behavioral Sciences

EWGT 2012

15th meeting of the EURO Working Group on Transportation

Retrieving real-time information to users in public transport networks: an application to the Lisbon bus system

David Alves*, Luis M. Martinez, José M. Viegas

Department of Civil Engineering, Instituto Superior Técnico, Lisbon Technical University, Avenida Rovisco Pais, Lisbon 1049-001, Portugal

Abstract

This paper presents a real-time trip-planner system for the public transport in Lisbon. This system has the capability of informing potential customers about which are the best routes to make the trip they want, when they want and what are the expected travel times, based on the actual locations of the public transport vehicles and the travel speeds that can be estimated for the various relevant road segments for the next hour. Using four months of operation log-files from the bus operator Carris, a process of data mining was created to analyse and classify the information of travel times and speeds. The trip-planner is built upon an agent-based model that aims to simulate the transport network operation and create a model to make short-term travel times forecast. A system of dynamic queries was introduced in order to evaluate the built model. The obtained results indicate that this tool, if deployed, could achieve high accuracy levels in predictions and become very useful and valuable for urban public transport users.

Keywords: Real-time data; transport demand estimation; public transport coordination; transfer matrix.

1. Introduction

One of the key challenges that the public transport sector faces is the reliability on the service schedules. According to (Lyons & Harman, 2002), the major grievances regarding public transportation are often delays in the arrival of buses and trains and the excessive time on board due to unforeseen events such as accidents or traffic. The static nature of most trip-planning systems prevents the real time assessment of travellers to information. Information as proved directly influence in riders' opinion towards public transport. Watkins (2011)

* Corresponding author.
E-mail address: dalves@civil.ist.utl.pt

 doi:10.1016/j.sbspro.2012.09.765

study found that for riders without real-time information, perceived wait time is greater than measured wait time which does not occur with riders using real-time information.

Within the context of intelligent transportation systems (ITS), last generation public transport networks are designed to collect, process and disseminate real time information. Research has focused in the last decade in developing solutions for real-time travel time estimation for private and public transportation systems, and also in the development of trip planner advisors applications. Yet, there is a gap on the developed research and the implemented models, which in the public transport sector remain not fully exploiting thee available ITS potentialities.

This paper introduces a new Intelligent Transport System (ITS) to support the system operation and upgrade the real-time information provided to users in order to increase the attractiveness and competitiveness of the public transport supply.

The proposed methodology encompasses a travel time estimation model for the different sections of a bus network, a simulation platform to generate the data required for the bus network operation, and a trip planner application based on real time information.

This tool will allow users to plan their immediately subsequent journeys through reliable information about the public transport supply, presenting the best options in terms of optimized route, optimized travel time and possible delays caused by accidents or incidents.

The formulated tool considers the availability of real-time exchange of data between a personal mobile device like a mobile phone, personal digital assistant (PDA), tablet or similar and the public transport network with the required data processing being remotely done by a central system.

This model was developed for the Lisbon bus and tram network, which operates 745 buses in 78 lines and 57 trams in 5 lines which account for an approximate average of 12,000 services per day.

After this brief introduction, this paper will present a summary of the literature review identifying the main methodologies being currently studied in order to predict travel times and build trip-planners. Afterwards, it will be discussed the framework of the model proposed in this paper followed by a description of the data used in this study, provided by the bus and tramway public transport operator in Lisbon Carris. Later, Section 5 presents a real-time prediction model for travel times in Lisbon's public bus and tramway network. Sections 6 and 7 present the trip-planner supportive methodology, a time-window adapted Dijkstra algorithm and the trip planner application integrated in a simulation model environment, using an Agent-based formulation. This paper is finalized with a test of the proposed model, some conclusions and future developments for this research.

2. System Framework

The model discussed in this article relies on an information platform (Data Centre) that filters and processes real-time information collected from GPS equipped buses. This data is then used to update the sections historical profiles and to improve real-time estimates by incorporating real-time information in a prediction algorithm. The travel time estimates of each section of the network are then used by a Trip-Planner application, which assesses with a time window adaptation of a Dijkstra's Shortest Path Algorithm the most suitable path for a given origin/destination and time of the day.

As seen in Figure 1, after being processed, the information is then broadcasted wirelessly to any mobile device (e.g. a smartphone).

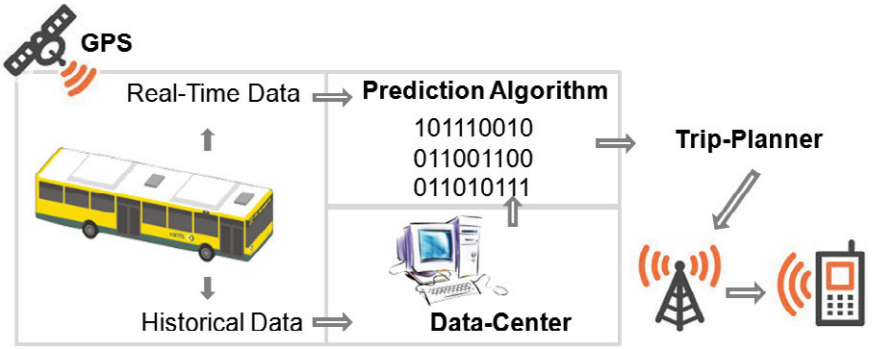

Figure 1 System Framework

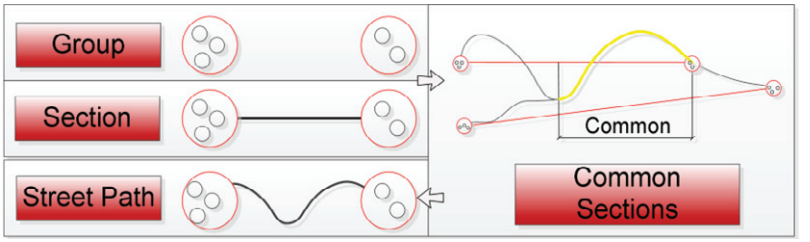

Figure 2 Characterisation of the different simulation elements

The spatial resolution used in the proposed model was constrained by the level of detail of the data provided, which only presented information about the time of stop of each vehicle during the system operation. The absence of information at the street level led to the consideration of a spatial construct that aimed to maximise the available information.

In order to conciliate the observed travel times between stops of multiple services, four spatial constructs were created: a group of stops which stand for stops located less than 30 meters apart in the network; a section that aggregates information between two groups of stops; a common section which represents the spatial overlap between two sections and the street path that represents the physical path between two consecutive stops of a service. The characterisation of these different elements is presented in Figure 2.

For the development of the current methodology, four different steps were undertaken: an initial data processing of historical data; the creation of a live travel time estimation model; a simulation environment to test a regular operation day of the Lisbon network and a trip-planner algorithm to estimate a customisable optimal route in public transport. All these components are presented in the following section.

3. Historical data processing

The historical data used to develop this model was based on log-files of four months of the operation of Carris (December 2009, January, April and May 2010 referring to the last but one network restructuring). Each file presented approximately 90,000 entries for weekend days and 300,000 entries for week days. These data files are

relative to the recorded time upon arrival at each stop of all the equipped vehicles of Carris. This data does not contain information about the time spent at each stop, just measuring the inter-stops time. This recorded time includes the halted time at the stop and the time to accelerate to the cruising speed at the origin and the deceleration time at the destination.

The files merged information about buses and trams that operated in Lisbon. Since there are only 5 tram lines operating in Lisbon, this study does not differentiate the bus and tram log-files, which is considered not to bias significantly the obtained results. From the available raw data, some additional variables were computed as the group to each stop belongs and the linear speed between consecutive stops.

A deeper analysis to the obtained data revealed the presence of outliers, mainly from GPS measuring errors or sporadic derivations of services. Every event observed less than ten times during the four months of operation were considered as outliers and removed from the data set. Additionally, speed measurements leading to values greater than the maximum travel speed (120 km/h in highways) were also considered as GPS location errors.

In order to generate the required inputs for the travel time prediction model, a speed profile analysis of the different sections of the city was carried out. For this purpose, the operational day was divided in 5 minutes periods (144 intervals) and computed the percentiles of speed for each section.

To characterise the linear speed of the constituent sections of the surface public transport network, a speed profile analysis was performed. The statistical distribution of speed data at different city areas and day periods was performed through the analysis of the percentiles of the available speed variable. In order to reduce the amount of data to process, four notable percentiles were selected to represent the shape of the probability density functions. This percentiles were: the first quartile ($P(x<X)=0.25$), the second quartile or median ($P(x<X)=0.5$), the third quartile ($P(x<X)=0.75$) and an upper limit lower than the fourth quartile, which intended to avoid the inclusion of outliers close to the observed maximum values. This percentile was set as $P(x<X)=0.9$ derived from a thorough analysis of the data, leading to more stable upper limit values of the speed. A global assessment of these speed profiles for the whole city is presented in Figure 3, where it can be observed decrease in speed during the peak periods and a steady behaviour between peaks.

In order to refine this analysis and increase the information available to use as input in the travel time prediction model, it was developed a cluster analysis to aggregate into groups of sections with different speed profiles during the operational period.

The agglomerative *Ward's Method* was considered as the most suitable option for the data available due to its ability to maximise the within group homogeneity and leading to the formation of groups with a significant number of members. The selection of the most adequate number of clusters was performed using the *Elbow's Method* (Ketchen & Shook, 1996), being 5 the number of groups that seemed more suitable. The general attributes of these clusters are presented in Table 1.

The analysis of the clusters speed profiles has shown a greater variation of the traffic average speed when approaching rush hours, where traffic congestion highly constrains the bus circulation. The structural network of the city lies mostly in Cluster 1, where it can be observed higher circulation speeds. Moreover, the analysis allowed grouping streets with unstable speed profiles (Cluster 2), where local incidents drastically affect the circulation speed, service efficiency and reliability. Another interesting result was the formation of a cluster that gathered large number of roads with dedicated bus lanes (Cluster 3). Contrary to what was expected, this cluster does not present an average circulation speed significantly higher than the average of the network (see Table 1).

Although outside the scope of this article, a more detailed analysis of this data has the potential to support improvement in route selection, suggesting criteria on how to avoid streets that cause service disruption due to traffic incidents or easy congestion.

Table 1 - Results of the cluster analysis

Cluster	N.º Elements	Av. Speed of percentile [0.5] [km/h]	Max. Speed [km/h]	Min. Speed [km/h]	St. Dev. Speed [km/h]
1	345	31.10	38.47	25.63	3.77
2	191	15.83	36.41	0.53	8.11
3	481	16.91	23.15	13.04	3.25
4	945	21.45	37.76	11.62	3.71
5	818	11.74	20.65	7.66	1.82

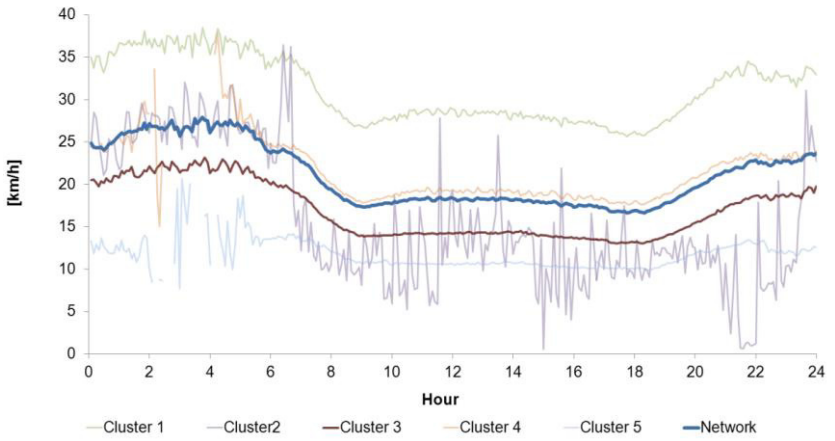

Figure 3 Clusters speed profiles (percentile 0.5)

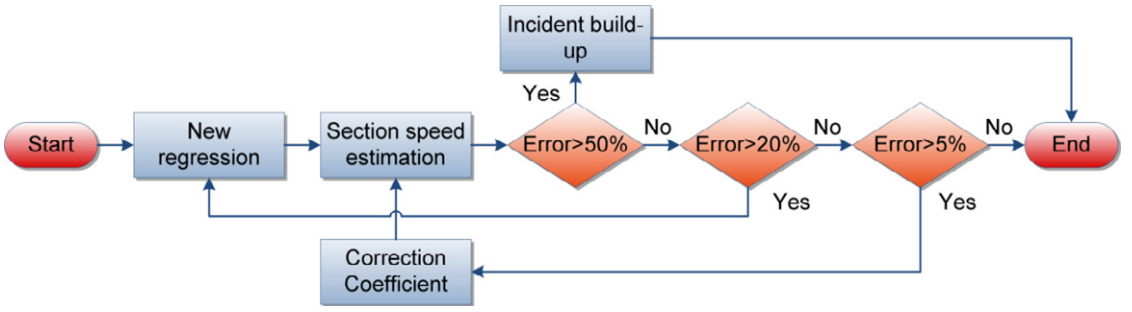

Figure 4 Model decision tree

4. Travel time prediction model

In order to estimate the time that buses take to perform each section of their routes, it was created a prediction system based on historical data and information in real-time. This prediction model is embedded in a decision tree that includes several methods to calculate the estimates for the next time intervals depending on the previous interval estimate of the relative error (see Figure 4).

The above mentioned model uses a 5 minutes time window and has as inputs for the regressions the historical speed profiles of the section and of the cluster and zone to which the section belongs. The real-time inputs are the instant speeds measured on the previous 30 minutes to the instant where the regression occurs.

On the first iteration, when the model does not have any estimate for travel time for the next intervals yet, a multivariate regression is performed. On the following intervals, the model assesses the estimated relative error and triggers different types of calibration procedures.

In this model, if the estimate of the last period does not satisfy the relative error threshold (5%), the model will correct the prediction. This threshold was set due to the consideration of the existence of measurement errors that could be greater than this deviation (GPS precision) and the instant speed variations that might not be captured in the available data. These corrections can trigger, depending on the level of relative error, three different functions: compute a new regression, make a correction to the regression estimates or trigger a build-up event for incidents. The other thresholds established for the relative error resulted from the observation of the case study, which may be highly dependent on the local context (i.e. level of congestion and presence of dedicated bus lanes). The established thresholds for these functions were:

- When the relative error is under 5%, the model preserves the estimates from the previous time period and projects the estimates for the next time periods;
- When the relative error is between 5% and 20%, the model computes a correction coefficient to the estimates to match the registered speed in the previous period and uses the same regression estimates with the correction coefficient to project for the next time periods;
- When the relative error is between 20% and 50%, a new regression of the model is triggered and the coefficients of each independent variable in the speed model are re-estimated;
- When the relative error is above 50%, the models triggers a build-up incident function, where the speed derivatives from the last time periods are used to predict speed reductions or incident solving in the next time periods.

The description of the different calibration models is presented next.

4.1. Multivariate linear regression

The multivariate linear regression was selected as the main methodology to estimate the travel speeds of sections for the next time periods that is frequently updated merging historical and real-time information. The selected procedure was formed by three groups of independent variables that try to explain the current travel speed of each section. These groups are:

- Sections historical data (4 percentiles);
- Zone historical data (4 percentiles);
- Recent information in the same Sections (last 5 periods of 5 minutes)

The sampling process for each section was designed to include in the estimate sections with similar characteristics of the current one. For this reason, were selected sections that belong to the same cluster (each

cluster being formed by sections with similar speed profiles along the day) and within the zone or neighbouring zones, which may present a similar traffic state. The sample sizes obtained for each regression vary between 40 and 200 elements with an average value of 82 cases.

The general equation is presented in:

$$Speed_t = b + \sum_{i=1}^{i=4} Speed_{Section_i}.a_i + \sum_{i=5}^{i=8} Speed_{Zone_i}.a_{4+i} + \sum_{h=1}^{h=5} Speed_{t-h}.a_{9+h} + \varepsilon \quad (1)$$

4.2. Correction coefficient

This procedure is activated when speed estimates require a small adjustment to fit the observed values. The correction coefficient is estimated as the ratio between the expected speed and the observed one, using the same regression parameters. This correction coefficient is then used, along with the regression coefficients, to predict speeds for the next 6 periods.

4.3. Incident Build-Up

The incident build-up procedure is also only triggered when speed estimates are significantly large. The developed formulation presents a simple formulation due to the lack of detailed incident information available from different sources (e.g. local authorities). On that case, the detection of incidents could be handled separately, not interfering with the regressions to be performed within the model.

The incident build-up model consists of an analysis of the behaviour of the speed derivative in presence of measured pronounced variations in traffic speed. When the observed deceleration reaches a threshold value the procedure launches a speed variation function, which is dependent on the derivate observed on the last N periods and on which the speed has been constantly decreasing or recovering. The estimate speed for the next time periods is then obtained with the following equation:

$$Speed_{t+1} = Speed_t + \frac{1}{N}\sum_{i=1}^{i=N} \frac{dSpeed_i}{dt} \quad (2)$$

When this procedure is triggered for the first time, it will require to be once again launched at least for the next two periods, or until the recovery of the normal situation that is assessed by comparing with the historical median of speeds in the section for a given time period.

5. Trip-Planner

5.1. Methodology

The model developed for the trip planner module was based on a Shortest Path Algorithm with Dynamic Time Windows (SPADTW), formulated upon a traditional Dijkstra algorithm. Dijkstra algorithm is a graph search algorithm that solves the single-source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree. This algorithm is often used in routing and as a subroutine in other graph algorithms.

In a public transport network defined by the service headways, the original formulation of Dijkstra's algorithm may no longer produce shorter paths because the axiom of separability of the optimal shortest path in optimal

sub-paths between intermediate nodes no longer applies. This is due to the fact that the shortest path to an intermediate node may correspond to using a direct service to reach that location, but implies a transfer to another service on the way to the end node. The use of a traditional formulation may lead to sub-optimal solutions for the destination node, whereas the shortest path to the end node may present disjoint shortest paths to an intermediate node. However, this problem may be solved by creating a synthetic network where each network node is decomposed in several multiple nodes, depending in the typology of the precedent and next node in the network. Furthermore, the current formulation considers that some arcs of the network may only be accessed in some specific time windows, which are set dynamically depending on the predicted arrival time of each bus at a specific stop. These time windows at each node are obtained through the travel time prediction model, allowing a very detailed characterisation of the availability of services for each time step. These adaptations to the node specification and node access availability produce an efficient SPADTW, which allows a fast and steady assessment of the best routing options at a given moment in time (Merrifield, 2004).

For this purpose, the current model encompasses all the required networks that will be traversed by the user (street network, transfer connections and bus network), allowing a detailed assessment of all the travel time components of the itinerary.

Yet, the current formulation may present the limitation of considering a compensatory objective function, which might lead to some solutions that present some features that are not easily accepted by users (i.e. excessive number of transfers). This problem may be surpassed by introducing in the objective function penalties relative to some attributes stated by the user that they wants to avoid (i.e. excessive number of transfers, excessive walking time).

The adopted model considers a generalised cost function (GC) which converts all the elements to time units. The trade-off coefficients between walking time (WT), transfer time (TT), number of transfers (NT) and on-vehicle time (OT) have to be established prior to the model computation. The general equation is:

$$GC = \alpha \cdot OT + \beta \cdot WT + \gamma \cdot TT + \delta \cdot NT \tag{3}$$

The obtained optimal path, for a given combination of the objective function parameters, is then converted to time and communicated to the user, presenting all the relevant information for his ride (total travel time, walking time and location of source stop, initial service estimated schedule, location and timing of the transfers, location of egress stop and walking path the destination).

The inputs required to compute the model are the location of the user and the desired destination. Taking into account the traffic conditions of the network, the model estimates all the expected bus arrivals times for the following 30 minutes at each stop (node) and computes the optimal path for the user at the moment of the query. If part of the path to be estimated is beyond a 30 minutes horizon, the model will use historical data as a predictor for service scheduling at those stops.

6. Public Transport network simulation model

In order to assess the potentiality and accuracy of the developed methodology, an Agent Based Model (ABM) was created. This model has two distinct operation modes: an offline and online mode that will be described in the next section. The general model framework is presented in Figure 5, where the different agents of the system are presented, showing their main activity in the model and their interactions thought the environment.

The model contains three main agents: the users, the buses and trams and the sections. The user's role in the model is to formulate a query to the system from an origin to a destination census block at a given time, and then execute the retrieved travel plan and compare the estimated with the experienced travel plan. The buses and trams operate the pre-established services with the available fleet, following the traffic conditions retrieved dynamically

by the environment. Finally, the sections contain the ability to gather historical and real-time information and predict their travel time every five minutes for the next 30 minutes period.

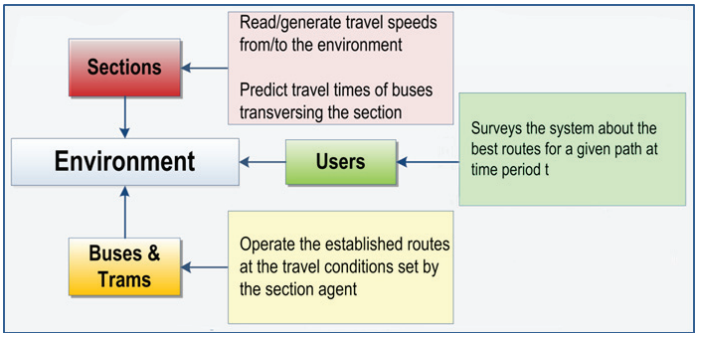

Figure 5 - Agent-based simulation model framework

The environment in the model is represented by six main components:

- The *section*, which is an abstract representation of the connection between groups of bus stops. This environment feature is simultaneously a component of the environment and an agent in terms of its ability to take decisions (encompassing the travel time prediction model).
- The *street paths*, which represent the real path travelled by buses while operating;
- The *common sections*, which stand for street segments that belong to the paths of different routes, being used as basis for the corresponding speed conciliation;
- The *stops and groups* of stops that represent the locations where Users board and egress;
- The *census blocks*, which are used as the reference to determine origins and destinations of u*sers*;
- The general *transport network,* encompassing all the above elements plus the p*edestrian network*, the *connectors* between the origins and destinations (census blocks) of *users* and the *transfers*, which represent the logic connection between the *pedestrian network* and the *stops*.

The next sections describe the different components and sub-models integrated in the ABM model that are required to simulate a regular day of operation of Carris.

6.1. Generation of traffic conditions in the network

This section describes how the above mentioned offline and online system modes operate. The first mode simulates, with no real-time information available, the traffic conditions on the different sections of the network, using the typical speed profiles with an additional random component to introduce variability. The second mode reads one Carris log-file directly and generates buses on the network following the operational records for that specific day Offline mode.

6.1.1. Offline mode

The developed offline mode creates synthetic operational days and can be used to test different traffic states, such as incidents or special events that might influence significantly the system operation. Furthermore, this

approach considers a perfect data acquisition status, which is far from the current situation where the system only receives data from vehicles when they reach a stop.

The procedure to generate travel times is based on a three components model:

- One relative to the impact of the historical data on the generation of the next period instant speed (12 variables);
- Another devoted to the last observed speeds in the section (6 variables);
- And a third component relative to a random variation of the instant speed. This component was modelled through the statistical distribution of the speeds observed in the sample. This random component at this stage of the study was based on the entire network speed profile and follows a normal distribution with an average speed of 23.01 km/h and a standard deviation of 5.13 km/h (1 variable).

The model contains in total 19 variables, whose weight in the final linear model is randomly generated for each period to ensure independency between consecutive speed estimates. Within each group the weights of each variable for the linear model vary from period to period. Yet, the weight of each group on the overall estimate is set as fixed. The real-time information will represent 50% of the instant speed estimation, the rest randomly split being by the other groups. The general equation is:

$$Speed(t) = \sum_{i=1}^{i=19} v_i . u_i \tag{4}$$

The used approach ensures significant travel time variability during the day, although considering perfect data availability, which may produce better model results than in a real context.

6.1.2. Online mode

The online mode reads directly a Carris log-file. This approach was mainly used to test the prediction model and the shortest path algorithm for a real operation day.

The system collects data from the buses or trams passages at stops, the Speeds at each section are estimated through a back propagation process. The final estimates, for each section, will result in a weighted contribution of each measurement, based on the common section and street path objects components. It is noteworthy that, since the system only receives data depending on vehicles real operation, the occurrence of new information may lead, in some sections with low bus frequency, to data unavailability for more than one time period (5 minutes).

The presented process results in an equation for each section based on the relation between street paths, common sections and sections. The travel time in a section can be then estimated by:

$$Speed(t) = \frac{d}{\frac{\sum_{i=1}^{i=N} travel\ time_i . \frac{d}{length_i} . P_i}{\sum_{i=1}^{i=N} P_i}} \tag{5}$$

7. Analysis of the results of the model

7.1. Offline travel time prediction model evaluation

In order to test the accuracy of the developed real-time prediction model, a comparison test between the estimates data median for a given path versus the developed model was undertaken.

The main difference, between the models is again the inclusion of a dynamic prediction algorithm that uses the travel times registered in the six 5 minute periods prior to the prediction instant and not only historical median values.

The results show (see Figure 6) for the real-time prediction model smaller deviation from real registered values when compared to the median model, presenting by higher frequencies of small deviations in estimates (less than 5 minutes) and lower frequencies of large errors (more than 15 minutes).

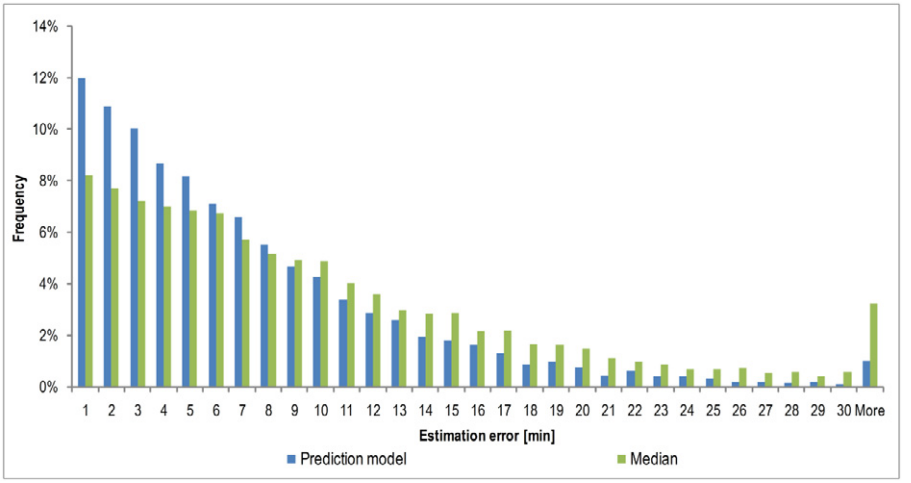

Figure 6 – Prediction model vs. Median speed profiles

The obtained results from the test illustrate the high improvement in accuracy when predicting travel times by incorporating real-time information in the prediction models.

7.2. Trip-planner test for a synthetic population of clients to measure the agenda adjustment

19 stops have been selected in order to perform a set of query tests to evaluate how the prediction model responds to requests on different places of the city and with different possible combinations of bus lines. A simple stop selection principle was defined: the stops should be homogenously distributed along the city and they should be located in easily accessible points by public transport (i.e. with multiple bus lines available).

In order to evaluate the reliability of the trip-planner, five indicators were assessed for the 3,240 tested scenarios.

As presented in Table 2, the observed relative error of the estimates is rather small (1.4%). Although, this value tends to increase with the length of the connection, the error propagation seems to be not significant. As in the previous indicator, the correlation of the estimates and real values is rather high.

Table 2 - Test indicators

Indicator	Observed Value
Average and Std. Dev. of the relative error	1.4 / 1.87 %
Correlation coefficient	0.99
Average and Std. Dev. on the time spent on transfers	0.33/0.92 min
Average number of transfers	1.07
Average and Std. Dev. on walking time	10.09 min / 12.06 min

The predicted required time spent in transfers seems to be accurately estimated, with deviations smaller than 0.92 min. The number of transfers observed between origins and destinations varies significantly along the day

and between the O/D pairs. This indicator is largely dependent on Carris network design, given priority to direct connections to some points sin Lisbon. Although the algorithm is not able to solve the quality of the connection between zones, it can significantly improve the level of service of these connections by minimizing the time lost in walking to/from stops and waiting. In terms of walking, the obtained solutions seem to find balanced walking times at the origins and at destination with the exception of trip extremes located close to each other (i.e. Campo de Ourique – Prazeres).

7.2.1. Comparison with offline data from Transporlis

A test was developed in order to compare the plans obtained with the trip-planner and the offline historical data based website from Transporlis. A preselected set of itineraries were tested and the results are summarized in Table 3, where is clear a significant difference in estimates of total travel time and waiting times on the transfers. Green values are estimates with a difference less than 20%, orange 20%-50% and red more than 50%. It should be noticed the travel between Belém and Campo Pequeno where it was suggested the same itinerary but with travel time predictions differing more than 50%. This is probably due underestimation of on-board times made by Transporlis website.

Table 3 Comparison example of the model with the Transporlis website estimate

	Origin	Dest.	Start Time	Duration [min]	Lines	Walk Origin [min]	Wait Origin [min]	Wait at transfers [min]	Walk Dest. [min]	Total on-board [min]
Transporlis	Oriente	B. Alto	20:00	56	794	0	10	0	7	39
	Graça	Calvário	12:00	29	28E,732	3	3	(5)+9	2	7
	Belém	C.Pequeno	18:00	39	15E,732	2	5	6	6	20
	Telheiras	C.Ourique	10:00	70	747,701	3	3	10	4	50
	P.Espanha	Alvalade	16:00	37	746,755	2	6	(3)+5	1	20
Trip Planner	Oriente	B. Alto	20:00	49	28,79	0	6	(2)+4	3	33
	Graça	Calvário	12:00	53	34,12	1	17	2	3	30
	Belém	C.Peque.	18:00	76	15E,732	1	7	1	4	53
	Telheiras	C.Ourique	10:00	77	747,701	5	14	5	9	44
	P.Espanha	Alvalade	16:00	51	746,44	1	6	(1)+7	3	32

8. Conclusions

This paper presents the formulation of a new Trip-Planner tool for the bus and tram system of the city of Lisbon. An extensive review showed intensive research on the development of travel time prediction algorithms, but a significant lack of tools that integrate a travel time estimation method along with a Trip planner component.

This work represents a first step on the development of this tool, presenting the several steps required for the creation of a robust system, ranging from the data collection and processing, the prediction dynamic travel times of bus routes in complex urban environments and the creation of a platform to communicate with the user.

The developed prediction model showed good results using a decision tree model which triggers a linear regression component that dynamically recalibrates the travel time estimates for each section for the next 30 minutes based on historical and real time information of section with similar speed profiles and geographical proximity.

In order to test all the system components, an agent based simulation model was built, aiming to recreate a regular day of the operation of the system. This model may allow the construction of different operation settings, as well as road network behaviours that can affect the services operation and evaluate of the performance of all the system components and the reliability and robustness of the information retrieved to users.

The obtained results from the model are very positive when compared with a synthetic speed model. The hypotheses that may still limit the accuracy of the online model are the low number of registers, in the short term,

of bus passages, which significantly limits and bias the regression model. Furthermore, the unit of analysis, not directly comparable to the one registered in the log file may also be a constraint to the ability to predict precisely the travel times. Without these limitations it is expected to have even better results.

A small test-bed example was then conduced to prove the value-added of this new formulation. For that purpose, a small set of notable points in the city were selected to assess their possible connections at different hours of the day, but with the same user attributes specification for the Dijkstra parameters (waking speed and willingness to accept an extra transfer). The obtained results are very promising, although a larger and more complex test to the model is required. Nonetheless, the information already retrieved by the model as well as the speed of computation of all the possible solutions, shows the great potential of this application for a future real world application in the city of Lisbon or in other cities around the world.

Acknowledgements

In this study it was used a comprehensive database formed by the log-file of Carris operation produced and owned by the company EFACEC. This data was obtained through a data availability protocol signed with the Transportation Focus Area of the MIT Portugal Program.

References

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research an analysis and critique. *Strategic Management Journal, 17*, 441-458.

Lyons, G., & Harman, R. (2002). The UK public transport industry and provision of multi-modal traveller information. *International Journal of Transport Management, 1*, 1-13.

Merrifield, T. (2004). *Heuristic Route Search in Public Transportation Networks*. Ohio University, Athens.

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice, 45*, 839-848.